CLAIMS

What is claimed is:

1. A method for controlling congestion at a network switch, the method comprising:

receiving a frame having a source identifier field corresponding to a source node and a destination identifier field corresponding to a destination node, the frame having been transmitted to the network switch through a first intermediate switch between the network switch and the source node;

characterizing traffic flow at the network switch; and

sending a first instruction from the network switch to the first intermediate switch to control traffic from the source node to the destination node.

2. The method of claim 1, wherein the first intermediate switch is an edge switch coupled to the source node.

3. The method of claim 2, wherein the first instruction sent to the first intermediate switch comprises an edge quench frame.

4. The method of claim 3, wherein the edge quench frame has a source identifier field corresponding to the destination node and a destination identifier field corresponding to the source node.

5. The method of claim 4, wherein the edge quench frame includes network switch congestion information.

6. The method of claim 5, wherein the edge quench frame includes network switch queue level information.

7. The method of claim 6, wherein the edge quench frame directs the first intermediate switch to control the allowed rate for transmitting from the source node and the destination node by half.

8. The method of claim 7, wherein the first intermediate switch and the network switch are connected using fibre channel.

9. The method of claim 1, wherein the frame was transmitted through a second intermediate switch between the first intermediate switch and the network switch.

10. The method of claim 9, further comprising:

sending a second instruction from the network to the second intermediate switch to control traffic from the source node to the destination node.

11. The method of claim 10, wherein the first instruction sent to the first intermediate switch comprises a path quench frame.

12. The method of claim 11, wherein the second instruction sent to the second intermediate switch comprises the path quench frame.

13. The method of claim 12, wherein the path quench frame has a source identifier field corresponding to the destination node and a destination identifier field corresponding to the source node.

14. The method of claim 13, wherein the path quench frame includes network switch congestion information.

15. The method of claim 14, wherein the path quench frame includes network switch queue level information.

16. The method of claim 15, wherein the path quench frame directs the first and second intermediate switches to reduce the allowed rate for transmitting from the source node and the destination node to 0bps.

17. The method of claim 1, wherein characterizing traffic flow comprises checking the network switch queue level.

18. The method of claim 17, wherein characterizing traffic flow comprises determining whether to transmit path quench or edge quench frames.

19. The method of claim 18, wherein path quench frames are transmitted when the queue level exceeds a high threshold.

20. The method of claim 19, wherein edge quench frames are transmitted when the queue level is between a high threshold and a low threshold.

21. The method of claim 20, wherein the edge quench and path quench frames include a buffer level indicator.

22. A method for controlling traffic flow between first and second end nodes through first and second intermediate nodes, the method comprising:

transmitting a first frame having a source identifier corresponding to the first end node and a destination identifier corresponding to the second end node, wherein the frame is transmitted at a first intermediate node to a second intermediate node between the first intermediate node and the second end node;

receiving a second frame from the second intermediate node, the second frame having a source identifier corresponding to the second end node and a destination identifier corresponding to the first end node, wherein the second frame includes instructions to adjust the current allowed rate from the first end node to the second end node; and

adjusting the current allowed rate from the first end node to the second end node upon receiving the second frame.

23. The method of claim 22, wherein the current allowed rate can not exceed the maximum allowed rate.

24. The method of claim 22, wherein adjusting the current allowed rate comprises:

determining that the second frame is an edge quench frame.

25. The method of claim 24, wherein the current allowed rate is adjusted after it is determined that the first intermediate node is an edge switch coupled to the first end node.

26. The method of claim 24, wherein the current allowed rate is adjusted after it is determined that the first intermediate node is coupled to a neighboring node that does not support congestion control.

27. The method of claim 25, wherein the first end node is a host.

28. The method of claim 27, wherein the second end node is storage.

29. The method of claim 25, wherein the first end node is storage.

30. The method of claim 29, wherein the second end node is a host.

31. The method of claim 25, wherein the current allowed rate is initially the maximum allowed rate.

32. The method of claim 31, wherein the current allowed rate is divided by two upon receiving an edge quench frame.

33. The method of claim 32, wherein the current allowed rate increases at a recovery rate.

34. The method of claim 33, wherein the recovery rate is dynamically set.

35. The method of claim 33, wherein the recovery rate is set based on information contained in the received edge quench frame.

36. The method of claim 35, wherein the recovery rate is set based on an input queue associated with the second intermediate node.

37. The method of claim 22, wherein adjusting the current allowed rate comprises:

determining that the second frame is a path quench frame.

38. The method of claim 37, wherein the current allowed rate is initially the maximum allowed rate.

39. The method of claim 38, wherein the current allowed rate is reduced to 0 bps upon receiving an path quench frame.

40. The method of claim 39, wherein the current allowed rate increases at a recovery rate.

41. The method of claim 40, wherein the recovery rate is dynamically set.

42. The method of claim 40, wherein the recovery rate is set based on information contained in the received path quench frame.

43. The method of claim 42, wherein the recovery rate is set based on an input queue associated with the second intermediate node.

44. A switch for controlling the traffic flow between a source node and a destination node, the switch comprising:

a first port for coupling to a first external node;

a second port for coupling to a second external node;

a first queue associated with the first port for receiving data from the first external node, the first queue including a first portion for holding data for transmission through the first port and a second portion for holding data for transmission through the second port; and

a filter coupled to the first queue, the filter configured to receive data from the first queue and determine whether transmission of the data should be delayed based on information received from the second external node.

45. The switch of claim 44, further comprising a filter queues, wherein the filter queues are configured to hold data set for delayed transmission.

46. The switch of claim 45, wherein each filter queue is associated with a flow.

47. The switch of claim 46, wherein the flow is traffic from a source node to a destination node.

48. The switch of claim 47, wherein the first queue is a virtual output queue.

49. The switch of claim 47, wherein each filter queue is associated with a priority.12

50. The switch of claim 49, wherein each filter queue is associated with an input port and an output port.

51. The switch of claim 44, further comprising a rate limiter coupled to a filter queue.

52. The switch of claim 51, wherein the amount of delay is determined by the rate limiter.

53. The switch of claim 52, wherein the rate limiter uses token buckets.

54. The switch of claim 53, wherein the amount of delay is determined based on information received from the second external node.

55. The switch of claim 54, wherein the number of tokens allocated to a filter queue associated with a flow is halved upon receipt of an edge quench frame from the second external node identifying the flow.

56. The switch of claim 55, wherein the number of tokens allocated to the filter queue associated with the flow increases at a recovery rate.

57. The switch of claim 56, wherein the recovery rate is dynamically determined.

58. The switch of claim 56, wherein the recovery rate is set based on second external node queue level information.

59. The switch of claim 54, wherein the number of tokens allocated to a filter queue associated with a particular flow is set to zero upon receipt of a path quench frame from the second external node identifying the particular flow.

60. The switch of claim 59, wherein the number of tokens allocated to the filter queue associated with the flow increases at a recovery rate.

61. The switch of claim 60, wherein the recovery rate is dynamically determined.

62. The switch of claim 60, wherein the recovery rate is set based on second external node queue level information.

63. An apparatus for controlling congestion, the method comprising:

means for receiving a frame having a source identifier field corresponding to a source node and a destination identifier field corresponding to a destination node, the frame having been transmitted to the network switch through a first intermediate switch between the network switch and the source node;

means for characterizing traffic flow at the network switch; and means for sending a first instruction from the network switch to the first intermediate switch to control traffic from the source node to the destination node.

64. The apparatus of claim 63, wherein the first intermediate switch is an edge switch coupled to the source node.

65. The apparatus of claim 64, wherein the first instruction sent to the first intermediate switch comprises an edge quench frame.

66. The apparatus of claim 65, wherein the edge quench frame has a source identifier field corresponding to the destination node and a destination identifier field corresponding to the source node.

67. A computer readable medium for controlling congestion, the computer readable medium comprising:

computer code for receiving a frame having a source identifier field corresponding to a source node and a destination identifier field corresponding to a destination node, the frame having been transmitted to the network switch through a first intermediate switch between the network switch and the source node;

computer code for characterizing traffic flow at the network switch; and

computer code for sending a first instruction from the network switch to the first intermediate switch to control traffic from the source node to the destination node.

68. The computer readable medium of claim 67, wherein the first intermediate switch is an edge switch coupled to the source node.

69. The computer readable medium of claim 68, wherein the first instruction sent to the first intermediate switch comprises an edge quench frame.